**Name – Pankaj Chawla**

**Roll No. - 001811001052**

**BE IT 4th Year Machine Learning Assignment 1**

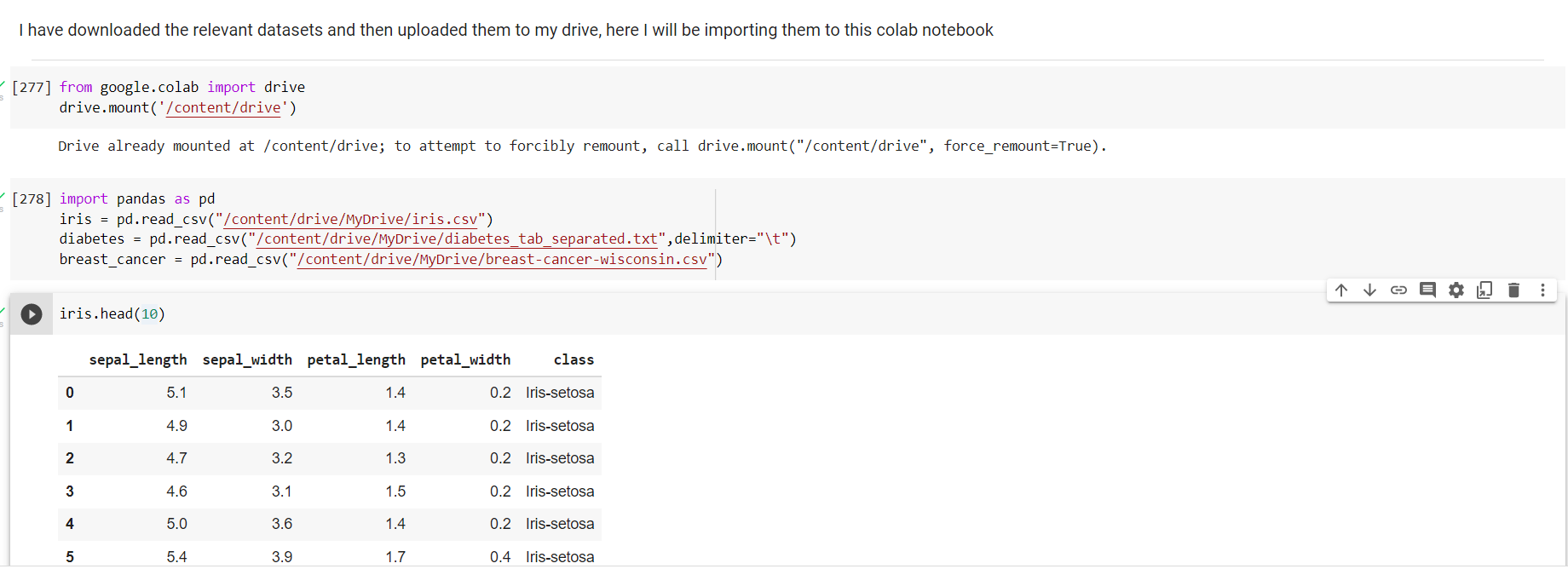
All steps in Google Collab – Google\_Collab

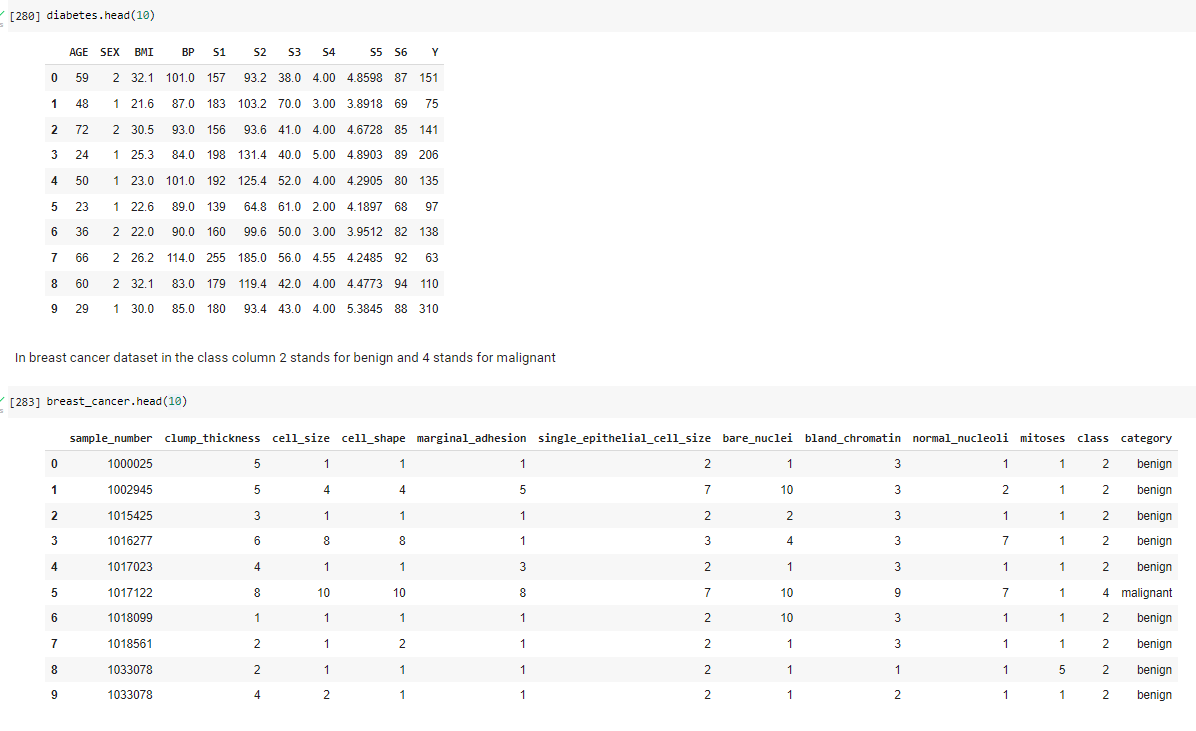
Procedure:-

* I first loaded the datasets from the CSV files available in the links and analyzed a bit of the data using pandas
* Then I imported all the relevant libraries that I was going to use for the assignment
* All three datasets were dealt with in the following manner:-
* First we used Naive Bayes Classification (Gaussian, Multinomial, Bernoulli). Bernoulli gave the worst performance both with and without parameter tuning
* Then we used Decision Tree Classifier (both gini and entropy) for both with and without parameter tuning.
* All performance metrics such as recall score, F1 score, precision score, and accuracy score were measured. I had really good results with the breast cancer dataset and the Iris Dataset with > 90% accuracy but the diabetes dataset had a dismal picture to it with ~60% accuracy on determining the gender attribute.
* Some sample screenshots are attached below for reference. As stated the mail already contains the decision tree images generated. However, since the notebook is pretty large I am unable to attach all the screenshots.

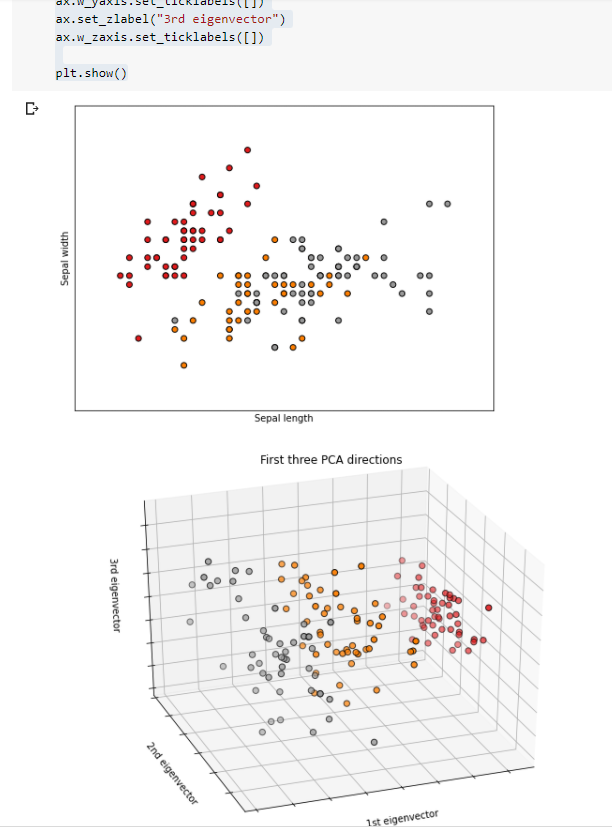
Screenshots:-

Loading data and importing libraries

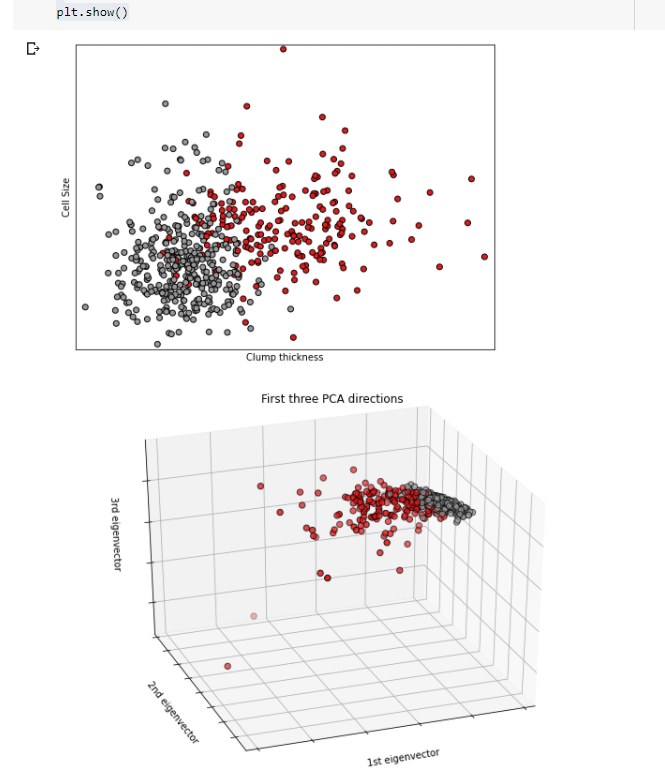




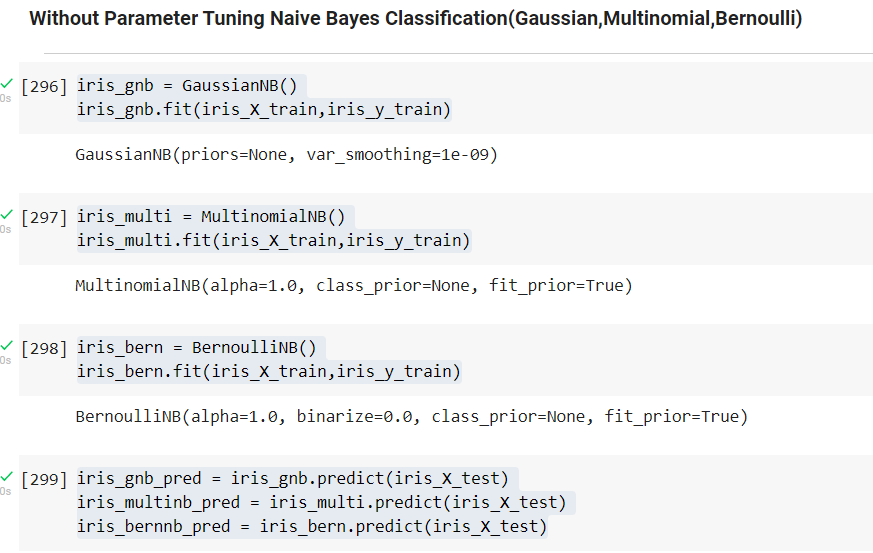
Iris dataset analysis:-



Breast Cancer dataset analysis:-



Construction of models:-





Error Metrics:-

